DOCUMENT RESUME

ED 476 864 TM 034 958

AUTHOR	Levy, Roy; Hancock, Gregory R.
TITLE	A Statistical Test for Comparing Nonnested Covariance Structure Models.
PUB DATE	2003-04-00
NOTE	26p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Comparative Analysis; *Statistical Analysis
IDENTIFIERS	*Covariance Structure Models; *Nested Data

ABSTRACT

While statistical procedures are well known for comparing hierarchically related (nested) covariance structure models, statistical tests for comparing nonhierarchically related (nonnested) models have proven more elusive. While isolated attempts have been made, none exists within the commonly used maximum likelihood estimation framework, thereby compromising these methods' accessibility and general applicability. The current work builds on a distance measure originally proposed by C. Rao (1945; 1949), and its application to distances between covariance structure models (A. Kumar and S. Sharma, 1999), thereby proposing a method for conducting a statistical test of such distances in order to assess formally the distinctness between models-nested or nonnested. An illustration is presented, and simulation evidence is provided to validate the performance of the proposed method. Two appendixes contain an illustration of the model for data generation and a program to compute distances between covariance matrices. (Contains 1 table, 2 figures, and 28 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 476 864

Running head: Test For Nonnested Models

A Statistical Test For Comparing Nonnested Covariance Structure Models

Roy Levy and Gregory R. Hancock

University of Maryland, College Park

[Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL, April 2003]

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

R. Levy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

TM034958

Correspondence concerning the paper should be directed to the first author at:

1230 Benjamin Building
University of Maryland
College Park, MD 20742-1115
levyr@wam.umd.edu

BEST COPY AVAILABLE

Abstract

While statistical procedures are well-known for comparing hierarchically related (nested) covariance structure models, statistical tests for comparing non-hierarchically related (nonnested) models have proven more elusive. While isolated attempts have been made, none exists within the commonly-used maximum likelihood estimation framework, thereby compromising these methods' accessibility and general applicability. The current work builds upon a distance measure originally proposed by Rao (1945; 1949), and its application to distances between covariance structure models (Kumar & Sharma, 1999), thereby proposing a method for conducting a statistical test of such distances in order to assess formally the distinctness between models – nested or nonnested. An illustration is presented, and simulation evidence is provided to validate the performance of the proposed method.

A Statistical Test For Comparing Nonnested Covariance Structure Models

The comparison of competing models is a keen interest for researchers interested in structural equation models and related techniques, be they path analysis, confirmatory factor analysis (CFA), or other members of the structural equation modeling (SEM) family. A number of indices have been developed for evaluating models individually and are often employed in comparing alternative models. Model comparisons involving statistical tests have generally been limited to likelihood ratio tests (χ^2 difference tests) for models in a hierarchically related system. We define hierarchically related systems as consisting of at least one nested and one general model such that the nested model is a special case of the general model. If M_A is defined by a vector of parameters $\boldsymbol{\theta}_A = (\theta_1, \ldots, \theta_{m_A})$ with ν_A degrees of freedom and M_B is defined by a vector of parameters $\boldsymbol{\theta}_B = (\theta_1, \ldots, \theta_{m_B})$ with ν_B degrees of freedom then M_B is nested within M_A when $\boldsymbol{\theta}_B$ is a subset of $\boldsymbol{\theta}_A$, i.e., $m_B < m_A$.

While the aim of most goodness-of-fit statistical tests is to evaluate the null hypothesis that one particular model fits the data, often a more pragmatic concern is whether or not two models approximate the underlying population equally well (Golden, 2000). Along those lines, Kumar and Sharma (1999) argued that in acknowledging: (1) all models are approximations, (2) multiple models might fit the data identically, and (3) it is sometimes necessary to respecify existing models, researchers are led to questions of discrepancy among competing models. Investigating the relation(s) among competing models directly might offer valuable evidence for assessing model discrepancy, ultimately as an aide in model selection. The current work is not as concerned with the discrepancies between individual models and the data as with the discrepancy between the models themselves. Further, the approach we take is to assess the discrepancy between models directly, rather than indirectly via a comparison of their respective

discrepancies to a fixed point, i.e., the sample data. Comparing models that fit the data equally well via the indirect strategy will not necessarily reveal any discrepancy between the models; a strategy of direct comparison will. This important distinction is visually represented in Figure 1. In the center of the diagram is the data to which Models M_1, M_2, and M_3 are fit. The dashed lines between the data and the models represent data-model (mis)fit. The collinearity between the data, M_1, and M_3 reflects that M_3 is nested within M_1. In hierarchical systems, the distinction between comparing models directly or indirectly via reference to the individual models' fit to the data is moot. The discrepancy between M_1 and M_3 is completely captured by the difference between the data-model fit of each model; here, χ^2 difference tests (of individual data-model fit) serve to assess the discrepancy between models. This is not necessarily the case in nonhierarchical systems. M_1 and M_2 fit the data equally well; the discrepancy between them (the solid line) cannot be assessed using only the models' respective data-model fit. We therefore aim at a procedure that statistically tests the discrepancy between competing models directly. Establishing statistical that tests of discrepancy for non-hierarchically related models has proven most challenging, though procedures aimed at this problem have been proposed. Still, no dominant paradigm exists for the comparison of nonnested models within the maximum likelihood estimation (MLE) framework.

In the next section of this paper, strategies for model comparison in various research settings are reviewed. Next, a distance measure for evaluating the differences between models originally proposed by Rao (1945; 1949) and first applied by Kumar and Sharma (1999) to distances between covariance structure models will be described. An empirical strategy for conducting a statistical test of such distances will be explored, the purpose being to derive a general test for models within the SEM framework that provides a statistical assessment of the

The difference between the χ^2 values for hierarchically related models M_3 and M_1, $\Delta\chi^2 = \chi_3^2 - \chi_1^2$, can be treated as a χ^2 with degrees of freedom equal to the difference between the number of free parameters in each model, $\Delta\nu = \nu_3 - \nu_1$. Under a true null hypothesis stating correct and equivalent data-model fit, $\Delta\chi^2$ follows a central χ^2 distribution with $\Delta\nu$; if the null hypothesis is false, $\Delta\chi^2$ follows the appropriate noncentral distribution. Such χ^2 techniques are also employed in evaluating model fit individually with respect to the observed data (i.e., not in a comparative setting as is the main concern of this paper). That is, any individual model can be thought of as a constrained version of a saturated model, i.e., a model that fits perfectly and has no degrees of freedom; thus the statistical test of the model χ^2 can be thought of as a test between a nested model and a general (saturated) model.

When models are not hierarchically related, the likelihood ratio test is no longer valid, and thus one typically resorts to a comparison of existing fit indices. As noted by several authors (e.g., Bagozzi & Yi, 1990; Byrne & Goffin, 1993; Widaman, 1985), differences in fit indices (absolute, parsimonious, and/or incremental) can provide a useful practical heuristic for assessing differences in data-model fit. However, these do not provide a formal statistical test. The Akaike Information Criterion, AIC (Akaike, 1987), is also commonly recommended for facilitating a choice among models, nested or nonnested. In short, the index allows for a determination of the relative replicability of the models under comparison, without providing a direct assessment of their overall adequacy of data-model fit. The AIC is a parsimonious fit index, in which absolute data-model fit is penalized for model complexity (i.e., having more parameters). Though common, such a strategy is not without criticism. First, as noted by Golden (2000) and Mulaik (2001), with increasing sample size such complexity penalties have less influence on the index (for a detailed discussion of this issue and subsequent criticisms of

the AIC, see Mulaik, 2001). There exist alternate information criteria, such as the Schwarz Information Criterion (SIC; Schwarz, 1978) and the Consistent AIC (CAIC; Bozdogan, 1987), which incorporate sample size into the penalty term. However such adjustments do not overcome the second and, for the purposes of this work, the most important shortcoming of this approach, namely that comparison of fit indices or information criteria does not constitute a formal statistical test of the difference in data-model fit of competing models, nested or nonnested.

To such end, attempts have been made on a number of fronts for statistically comparing nonnested models. Work has progressed in combining the psychometric literature on estimating latent variables with the econometric literature on testing nonnested models estimated with instrumental variables. Oczkowski and Farrell (1998) implemented a two-stage least squares (2SLS) estimator derived from Bollen (1996) and conduct a Cox-type *J*-test and an encompassing *F*-test derived from Smith (1992) to compare models employing different measures of market orientation. Unfortunately, as Oczkowski and Farrell (1998) noted, simulations suggest the *J*-test tends to over-reject the true null hypothesis while the *F*-test tends to possess less power when the null is false.

Other approaches in the comparison of nonnested models have come from employing Bayesian analyses. Most Bayesian methods of comparing nonnested models provide posterior probabilities that are roughly interpretable as probabilities of model correctness, but only if the correct model is part of the specified set. A Bayesian Cross Validated Likelihood method (BCVL) originally proposed by Rust and Schmittlein (1985) has been shown to overcome this qualification, though if the correct model is not among those being compared, posteriors can only be treated as conditional probabilities (Fornell & Rust, 1989). The BCVL method consists of

specifying a prior probability of correctness for each competing model, estimating each model from half of the observed data and calculating posterior probabilities for each model by combining (via Bayes theorem) the prior probability for the model and the likelihood of observing the second half of the data. There have been several extensions of the BCVL since it was first proposed, notably generalizations to allow the comparison of nonnested models (Fornell & Rust, 1989) and the comparison of nonnested SEM models that may have unique observable indicators (Rust et al., 1995)

As potentially useful as the BCVL method and its extensions may be for model selection, there is no statistical test for it to compare models. To conduct a pseudo-hypothesis test, Rust and Schmittlein (1985) advocate setting priors for each model to mimic a hypothesis test. In comparing hierarchically related models (akin to a classical test at the .05 significance level), the nested model would have a prior of .95 and the rejection of the null hypothesis of equivalent data-model fit would be justified when the posterior probability for the general model exceeds .5 (Rust & Schmittlein, 1985; for a clear illustration of the relationship between the prior and model selection, see Fornell & Rust, 1989). However appropriate this approach may be for determining which model fits better, which was the concern of Rust and Schmittlein (1985) and Fornell and Rust (1989), it reiterates the difference between the focus of this work and others. It is possible for the pseudo-hypothesis test to indicate that two models do not differ in data-model fit, even when the models are different. This may occur because, like other methods that involve the likelihood, the BCVL indirectly compares the models by virtue of assessing the discrepancy of each to a fixed point, the sample data. When the likelihood of the data under each model is the same (e.g., M_1 and M_2 in Figure 1), comparing the model χ^2 values or computing the BCVL will

not be able to distinguish between models. The method proposed in this paper is aimed at being able to reveal model difference (when present) even when the models fit the data equally well.

Along the lines of a more classical approach, Golden (2000) extends the work of Vuong (1989), describing a large sample procedure designed to test if the estimated data-model fit for one model is significantly different from that of another model. Golden (2000) demonstrates this test on logistic regression models rather than covariance structure models, and while the general nature of the test suggests its applicability to other analyses warrants investigation, the procedure (1) is tied to residuals of observations, which are not the unit of analysis in covariance structure analyses, and (2) involves the assessing the discrepancy between models indirectly via their discrepancy to the data.

The development of statistical tests of nonnested models has been drawing attention from varying perspectives, as there is no standard paradigm for conducting such tests of nonnested models in SEM and related techniques. It is to that end that this enterprise is directed, namely, the derivation of a test of covariance structure models that is applicable under very general conditions. This alternative and potentially more accessible line of reasoning draws upon a distance measure originally proposed by Rao (1945; 1949). Following Kumar and Sharma's (1999) application of Rao's Distance to describe the distance between two covariance structure models, our work proposes a procedure for conducting hypothesis tests of covariance structure models that is applicable under general conditions.

Rao's Distance Measure

For two p-variate normal distributions, each has p means and $p(p+1)/2$ variances and covariances. The well-known Mahalanobis distance,

$$D^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (3)$$

(Mahalanobis, 1936), assesses the disparity between the two distributions' px1 mean vectors (centroids $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$) alone, thereby assuming homogeneity of covariance matrices. Rao (1945; 1949) extended this work to create a distance measure able to assess distributional disparity across all $p(p+3)/2$ first and second moments. If one wishes to assume identical mean vectors ($\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$), a simplified version of Rao's Distance (RD) may be used to assess the disparity between two p-variate distributions' covariance matrices. As shown by Atkinson and Mitchell (1981),

$$RD(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = \left\{ .5 \sum_{i=1}^{p} (\ln \lambda_i)^2 \right\}^{1/2} \quad (4)$$

where $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are the covariance matrices, and λ_i are the i=1,...,p eigenvalues that satisfy the equation

$$|\boldsymbol{\Sigma}_2 - \lambda_i \boldsymbol{\Sigma}_1| = 0 \quad (5)$$

As noted by Kumar and Sharma (1999), if $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are positive definite and symmetric, all the p eigenvalues will be strictly positive.

RD has several desirable characteristics. As discussed by Kumar and Sharma (1999), it satisfies standard properties of distance such as symmetry with respect to its arguments (i.e., $RD(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = RD(\boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_1)$) and invariance to changes in variables' scale when applied to both populations. In addition, as the reader may easily verify, RD is always nonnegative and the distance from any population to itself is zero.

Kumar and Sharma (1999) suggest the descriptive application of RD to the model-implied covariance matrices arising from competing models imposed upon the same sample covariance

matrix. The disparity between two models can be assessed using $RD(\hat{\Sigma}_1, \hat{\Sigma}_2)$ where $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are model-implied covariance matrices. The focus of this paper is to extend the use of RD as a comparative index by incorporating a statistical test of significance.

Rao (1945) provided a framework for conducting tests of statistical significance for the metric. In the case where one population is specified and the other is represented only by a sample, an estimate of the distance between the populations is defined as the distance between the set of maximum likelihood parameter estimates based on the sample and the set of parameters that define the specified population. Rao (1945) also provided a large sample approximation of the variance of the estimated distance. If $\boldsymbol{\theta}_0$ is a vector of parameters defining the null population and $\hat{\boldsymbol{\theta}}$ is a vector of the maximum likelihood estimates obtained from a sample, then the statistic

$$w = \frac{RD(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)}{\left[V\left(RD(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0)\right)\right]^{1/2}} \tag{6}$$

is asymptotically distributed as a standard normal variate (Rao, 1945). The extension to the case of two independent samples case is conceptually straightforward. RD is calculated between the two sets of maximum likelihood estimates and the standard error term is adjusted to reflect the variability associated with estimating the distance from two independent samples (Rao, 1945).

In terms of covariance structure modeling, the derivation of a sampling distribution for the distance between independent sample covariance matrices from the same population would rest (in part) on the distribution of sample covariance matrices. It is well-known that the variances and covariances obtained from randomly selecting data from a p-variate normally distributed population follow a Wishart distribution (Wishart, 1928). In the case of independent covariance matrices, it is possible that an estimate of the sampling distribution could be derived. Unlike

such a scenario, however, model-implied covariance matrices fit to a common sample covariance matrix are not independent. The analytical derivation of a standard error for *RD* requires (in part) the cross-group covariances of model-implied variances and covariances. Such derivations are generally unwieldy as well as model-dependent, thus making a derivation of the theoretical sampling distribution and the hypothesis testing it facilitates practically intractable. The current paper therefore offers an empirical approach to the estimation of the sampling distribution in question, thereby facilitating hypothesis testing of the disparity between models, nested or nonnested, applied to the same data.

An Empirical Approach

Consider Model 1 with m_1 parameters contained in set $\boldsymbol{\theta}_1 = (\theta_{11},\ldots,\theta_{1m_1})$ and Model 2 with m_2 parameters contained in set $\boldsymbol{\theta}_2 = (\theta_{21},\ldots,\theta_{2m_2})$. Model 1 and Model 2 may or may not be correct in truth; the test proposed concerns the distinction between them regardless of their correctness. Next, consider a Model C with m_C parameters contained in set $\boldsymbol{\theta}_C = (\theta_{C1},\ldots,\theta_{Cm_C}) = (\theta_{11},\ldots,\theta_{1m_1}) \cap (\theta_{21},\ldots,\theta_{2m_2})$. This model, referred to as the "Child" model, contains the complete set of parameters common to both models.

For any population covariance matrix fitting Model C, Model 1 and Model 2 also fit perfectly. In terms of *RD*, $RD(\boldsymbol{\Sigma}_C, \boldsymbol{\Sigma}_1) = RD(\boldsymbol{\Sigma}_C, \boldsymbol{\Sigma}_2) = RD(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = 0$, where $\boldsymbol{\Sigma}_j$ refers to the covariance matrix for the j^{th} model. Turning to samples, for any sample covariance matrix **S** drawn from a population that fits Model C, Model 1 and Model 2 will fit better, but only due to chance. $RD(\hat{\boldsymbol{\Sigma}}_C, \hat{\boldsymbol{\Sigma}}_1)$ and $RD(\hat{\boldsymbol{\Sigma}}_C, \hat{\boldsymbol{\Sigma}}_2)$ fluctuate above 0 only randomly, and thus $RD(\hat{\boldsymbol{\Sigma}}_1, \hat{\boldsymbol{\Sigma}}_2)$ also fluctuates randomly above 0. It is the distribution of this latter term, i.e., the

distribution of RD between model-implied covariance matrices of nonnested models under the null condition of equivalent population fit, which serves as the focus of the current work. In light of the aforementioned difficulties in analytically deriving the sampling distribution of $RD(\hat{\Sigma}_1, \hat{\Sigma}_2)$ under null conditions, a resampling strategy is employed to approximate this distribution and facilitate significance tests.

For *nxp* sample data matrix **Y** with covariance matrix **S**, the distance between Model 1 and Model 2 is

$$RD(\hat{\Sigma}_{1,\mathbf{Y}}, \hat{\Sigma}_{2,\mathbf{Y}}) = \left\{ .5\sum_{i=1}^{p} (\ln \lambda_i)^2 \right\}^{1/2} \tag{7}$$

where λ_i are the i=1,...,p eigenvalues that satisfy the equation

$$\left| \hat{\Sigma}_{2,\mathbf{Y}} - \lambda_i \hat{\Sigma}_{1,\mathbf{Y}} \right| = 0 \tag{8}$$

and $\hat{\Sigma}_{1,\mathbf{Y}}$ and $\hat{\Sigma}_{2,\mathbf{Y}}$ are the model-implied covariance matrices obtained from fitting Model 1 and Model 2 to the data in **Y**, respectively.[1] We now seek to derive a null sampling distribution against which $RD(\hat{\Sigma}_{1,\mathbf{Y}}, \hat{\Sigma}_{2,\mathbf{Y}})$ may be evaluated.

To that end, the model-implied covariance matrix $\hat{\Sigma}_{C,\mathbf{Y}}$ is obtained by applying Model C to **S**. The raw data in matrix **Y** may then be transformed (Bollen & Stine, 1993) to yield $\mathbf{Z} = \mathbf{Y}\mathbf{S}^{-1/2}\hat{\Sigma}_{C,\mathbf{Y}}^{1/2}$, where the covariance matrix of data in **Z** is now $\hat{\Sigma}_{C,\mathbf{Y}}$. For these data Model C fits perfectly, as do Models 1 and 2. That is, $RD(\hat{\Sigma}_{C,\mathbf{Z}}, \hat{\Sigma}_{1,\mathbf{Z}}) = RD(\hat{\Sigma}_{C,\mathbf{Z}}, \hat{\Sigma}_{2,\mathbf{Z}}) = RD(\hat{\Sigma}_{1,\mathbf{Z}}, \hat{\Sigma}_{2,\mathbf{Z}}) = 0$. For a pseudosample $\mathbf{Z}^*$ bootstrapped from **Z**, the RD value representing the distance between Model 1 and Model 2 will only differ

[1] As there are multiple models fit to data both prior to and after a transformation, a richer notation is required. In eqs. (7) and (8) and hereafter, model-implied covariance matrices will be indexed by two subscripts, the first denoting the model, the second denoting the data to which the model is fit.

from (i.e., be greater than) 0 randomly. Upon repeated bootstrap resampling from **Z**, a sampling distribution for these RD values can be derived empirically. This distribution is an estimate of sampling variability under the null condition that $RD(\hat{\Sigma}_1, \hat{\Sigma}_2) = 0$, and thus serves as the frame of reference for evaluating the observed $RD(\hat{\Sigma}_{1,Y}, \hat{\Sigma}_{2,Y})$ from the original raw data in **Y**. Specifically, for a desired α-level, if the observed $RD(\hat{\Sigma}_{1,Y}, \hat{\Sigma}_{2,Y})$ exceeds the $(1-\alpha)\times 100^{th}$ percentile value in the empirical sampling distribution, then the null hypothesis of model equivalence is rejected and the inference is made that Model 1 and Model 2 are different.

Illustration

Let Model G be the full model depicted in Figure 2; let Model 1 be the constrained version of Model G such that the cross loadings, λ_{41} and λ_{63} (the dashed paths) are constrained to be 0. Let Model 2 be the constrained version of Model G such that ϕ_{21} and ϕ_{12} (the bold bi-directional factor covariances) are 0. Model 1 and Model 2 are therefore nonnested models; the discrepancy between them (with respect to a common set of data) is of interest, and it is such a discrepancy that stands in need of formal statistical test.

A multivariate normal data set **Y** of size n = 200 cases was generated based upon Model G where the loadings ($\mathbf{\Lambda}$) for the measured variables and the covariance matrices for the factors ($\mathbf{\Phi}$) and errors ($\mathbf{\Theta}$) are given in Appendix A. Data was generated in GAUSS (Aptech Systems, 1996) by using Vale and Maurelli's (1983) extension of the method originally proposed by Fleishman (1978). Specifically, the program NNORMULT (Nevitt & Hancock, 1999) was used. Both Model 1 and Model 2 were fit to this data using EQS (Bentler, 1998), yielding

model-implied covariance matrices $\hat{\Sigma}_{1,\mathbf{Y}}$ and $\hat{\Sigma}_{2,\mathbf{Y}}$. Computation of $RD(\hat{\Sigma}_{1,\mathbf{Y}}, \hat{\Sigma}_{2,\mathbf{Y}})$ was performed in GAUSS.[2]

The Child Model containing the complete set of parameters common to both models is the constrained version of Model G where both sets of constraints have been imposed; λ_{41}, λ_{61}, ϕ_{21}, and ϕ_{32} are all constrained to equal 0. Model C was fit to the data, yielding model-implied covariance matrix $\hat{\Sigma}_{C,\mathbf{Y}}$. The raw data matrix **Y** was then transformed to yield $\mathbf{Z} = \mathbf{Y}\mathbf{S}^{-1/2}\hat{\Sigma}_{C,\mathbf{Y}}^{1/2}$, where **S** is the covariance matrix of the raw data **Y**, resulting in transformed data matrix **Z** with covariance matrix $\hat{\Sigma}_{C,\mathbf{Y}}$.

A total of 200 bootstrapped pseudosamples, $\mathbf{Z}_k^*$ $k = 1,\ldots,200$, each of size 200 were obtained by resampling with replacement from **Z**. For each bootstrapped sample, Model 1 and Model 2 were fit to the data and $RD(\hat{\Sigma}_{1,\mathbf{Z}_k^*}, \hat{\Sigma}_{2,\mathbf{Z}_k^*})$ was calculated. The resulting empirical distribution of $RD(\hat{\Sigma}_{1,\mathbf{Z}_k^*}, \hat{\Sigma}_{2,\mathbf{Z}_k^*})$ reflects the (sampling) variability in $RD(\hat{\Sigma}_1, \hat{\Sigma}_2)$ under the null condition of no discrepancy between models. If the observed $RD(\hat{\Sigma}_{1,\mathbf{Y}}, \hat{\Sigma}_{2,\mathbf{Y}})$ (eq. (7)) exceeds the $(1-\alpha)\times 100^{\text{th}}$ percentile value in this empirical sampling distribution, then the discrepancy between Model 1 and Model 2 is statistically significant.

In order to assess the performance of the bootstrap and the possible relationship between statistical tests of RD and $\Delta\chi^2$ in the case of nested models, Model C was fit to each bootstrapped pseudosample. Let

$$T_{j,k} = (n-1)(\ln\left|\hat{\Sigma}_{j,\mathbf{Z}_k^*}\right| - \ln\left|\mathbf{Z}_k^*\right| + tr[\mathbf{Z}_k^*\hat{\Sigma}_{j,\mathbf{Z}_k^*}^{-1}] - p) \qquad (9)$$

[2] A GAUSS (Aptech Systems, 1996) program for computing RD is given in Appendix B

be a test statistic associated with fitting the j^{th} model to p-variate bootstrapped data $\mathbf{Z}_k^*$. The difference between such test statistics for hierarchically related models (e.g., $\Delta T_k = T_{C,k} - T_{1,k}$ and $\Delta T_k = T_{C,k} - T_{2,k}$) approximates $\Delta\chi^2$.[3] Fitting Model C to the bootstrapped data in addition to fitting Model 1 and Model 2 facilitates an investigation of the relationship between RD and $\Delta\chi^2$ in the case of nested models (i.e., Model C is nested within both Model 1 and Model 2). For each bootstrapped pseudosample, both RD and ΔT were calculated for the distance between Model C and Model 1 and for the distance between Model C and Model 2. To evaluate the extent of the agreement between the proposed test of RD and the $\Delta\chi^2$ test, the process of data generation, fitting, transformation, bootstrapping, and fitting was repeated 100 times.

Results

Summary results from the calculation of the 95th percentile over the 100 replications of the empirical null distribution of $RD(\hat{\Sigma}_1, \hat{\Sigma}_2)$ are given in Table 1, though in the absence of an analytical expectation, they are neither easily interpretable nor of interest. In practice, the empirical 95th percentile from one set of bootstrapped samples would serve as our criterion for testing $RD(\hat{\Sigma}_{1,Y}, \hat{\Sigma}_{2,Y})$ at the .05 level. What is of interest is the relationship between the proposed test of RD and the $\Delta\chi^2$ test in the cases of hierarchically related models (i.e., Model 1

[3] We hesitate to formally treat any T_j (and, by extension, any ΔT) as χ^2 statistics for several reasons (Bollen & Stine, 1993). First, the distribution of T_j (and ΔT) to χ^2 in an approximate result. Second, while the data-generating population is continuous, the bootstrap distribution is necessarily discrete. Third, to treat T_j (or ΔT) as a χ^2 requires an assumption of multivariate normality, which, even if met in the data-generating population, does not necessarily hold in transformed data that is to be bootstrapped. Following Bollen and Stine (1993), we acknowledge these limitations and proceed to treat T_j and ΔT as approximations to the model χ^2 and $\Delta\chi^2$, respectively.

to Model C and Model 2 to Model C); if (the proposed test of) RD is operating in the same way as (the test of) $\Delta\chi^2$, there ought to be a monotonic relationship between the empirical null sampling distributions of RD and ΔT. Inspection of plots for both hierarchical systems indicated a nonlinear monotonic relationship between RD and ΔT. To assess the degree of agreement between RD and ΔT, Spearman's rank-order correlation coefficient (r_S) was computed between RD and ΔT for the 200 bootstrapped values within each replication, for each of the two nested comparisons. Over 100 replications, the average values of r_S for the comparison between Models 1 and 2 to Model C were 0.998323 and 0.999477, respectively, and the *worst* values of r_S for the comparison between Models 1 and 2 to Model C were .993817 and .998764, respectively (Table 1). These values indicates a high degree of monotonicity and strongly support our contention that that the result of the proposed test of RD would be consistent with the result of the $\Delta\chi^2$ test. That is, these results imply that the behavior of RD in the case of sampling from a true null condition almost perfectly mimics that of $\Delta\chi^2$ in the case of nested models.

Discussion

The concern of this investigation is the performance of the proposed procedure for testing RD for competing covariance structure models. In the case of nested models, the test of $\Delta\chi^2$ is the dominant paradigm within the framework of MLE of SEM models. It is argued here that the level of agreement between the proposed test of RD and the $\Delta\chi^2$ test is excellent. The observed Spearman rank-order correlations indicate a high degree of monotonicity between the distributions of RD and the $\Delta\chi^2$; we have strong evidence to believe that, for any significance level, the results of the proposed test of RD (i.e., reject or retain the null hypothesis of no model

discrepancy) would be just that of the $\Delta\chi^2$ test. In the case of nested models, the proposed test of RD meets the standard of consistency with the $\Delta\chi^2$ test. There is no such standard for evaluating the performance of the empirical distribution of RD in the case of nonnested models. Its justification rests on appropriateness of employing RD in the analysis of covariance structure models (Kumar & Sharma, 1999) and the logic of the procedure presented here. If the logic of the process is sound, the proposed procedure ought to mimic well-known procedures in the case of nested models. Simulation evidence presented above suggests that this is indeed so. Further, the method proposed does not rest on distributional assumptions of the metric; the distribution is empirically generated, allowing the extension to the case of comparing nonnested models.

Conclusion

Though work has progressed on several fronts in the statistical analysis of nonnested covariance structure models, there is currently no standard procedure for the testing of differences between such models. A distance measure conceived by Rao (1945; 1949), extended by Atkinson and Mitchell (1981) and previously applied to covariance analysis (Kumar & Sharma, 1999) was described. This work builds upon the application of this measure to covariance structure modeling; we propose a method for conducting a statistical test of the measure in order to formally assess the distinctness between models – nested or nonnested. An empirical approach for the assessment of the discrepancy between the fit of competing Models 1 and 2 was described and is summarized as:

(1) For sample data matrix **Y** with covariance matrix **S**, fit Model 1 and Model 2 to yield model-implied covariance matrices $\hat{\Sigma}_{1,Y}$ and $\hat{\Sigma}_{2,Y}$

(2) Compute $RD(\hat{\Sigma}_{1,Y}, \hat{\Sigma}_{2,Y})$

(3) Define the Child Model as $\theta_C = (\theta_{C1},\ldots,\theta_{Cm_C}) = (\theta_{11},\ldots,\theta_{1m_1}) \cap (\theta_{21},\ldots,\theta_{2m_2})$

(4) Fit the Child Model to $\mathbf{Y}$ to yield model-implied covariance matrix $\hat{\Sigma}_{C,\mathbf{Y}}$.

(5) Transform the data (Bollen & Stine, 1993) to yield $\mathbf{Z} = \mathbf{Y}\mathbf{S}^{-1/2}\hat{\Sigma}_{C,\mathbf{Y}}^{1/2}$

(6) For $k = 1,\ldots,200$

   a. Bootstrap the k^{th} pseudosample from $\mathbf{Z}$, $\mathbf{Z}_k^*$

   b. Fit Model 1 and Model 2 to $\mathbf{Z}_k^*$, yielding $\hat{\Sigma}_{1,\mathbf{Z}_k^*}$ and $\hat{\Sigma}_{2,\mathbf{Z}_k^*}$

   c. Compute $RD(\hat{\Sigma}_{1,\mathbf{Z}_k^*}, \hat{\Sigma}_{2,\mathbf{Z}_k^*})$

(7) If the observed $RD(\hat{\Sigma}_{1,\mathbf{Y}}, \hat{\Sigma}_{2,\mathbf{Y}})$ from (2) exceeds the $(1-\alpha)\times 100^{th}$ percentile value from those compiled from (6), reject the null hypothesis of model equivalence.

An illustration designed to mimic applied analyses was embedded in a simulation designed to assess the approach. The results of these investigations are quite promising; the proposed procedure is in accordance with existing procedures for the comparison of nested models. It is argued that employing the procedure in the comparison of nonnested models is justified on logical grounds.

Further research is currently being conducted to investigate the proposed procedure. In the context of the relationship between RD and $\Delta\chi^2$ in the case of nested models, areas of interest are (1) the association between p-values obtained from the test of RD and the test of $\Delta\chi^2$ and (2) the analytical relationship between RD and $\Delta\chi^2$. Other avenues of pursuit include investigating the impact of both sample size and the number of bootstraps, and extensions to multiple model comparisons. The Child Model has been defined as the model with the set of parameters common to the competing models; another line of research might include the viability of other definitions, e.g., a subset of the parameters common to the competing models. Further, the discussion of RD presented here concerns models in which mean vectors are assumed equal. An equation for RD has been derived in the case of dissimilar mean vectors (Atkinson & Mitchell, 1981); the method proposed here might then be applied to the comparison of mean structured models with augmented moment matrices (Sörbom, 1974).

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317-332.

Aptech Systems. (1996). *GAUSS System and Graphics Manual*. Maple Valley, WA: Aptech Systems, Inc.

Atkinson, C., & Mitchell, A. F. S. (1981). Rao's distance measure. *Sankhya, 43,* 345-365.

Bagozzi, R. P., & Yi, Y. (1990). Assessing method variance in multitrait-multimethod matrices: The case of self-reported affect and perceptions at work. *Journal of Applied Psychology, 75,* 547-560.

Bentler, P. M. (1998). *EQS Structural Equations Program*. Encino, CA: Multivariate Software, Inc.

Bollen, K. A. (1996). An alternative two stage least squares (2sls) estimator for latent variable equations. *Psychometrika, 61,* 109-121.

Bollen, K. A., & Stine, R. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen & J.S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: SAGE Publications.

Bozdogan, H. (1987). Model-selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52,* 345-370.

Byrne, B. M., & Goffin, R. D. (1993). Modeling MTMM data from additive and multiplicative covariance structures: An audit of construct validity concordance. *Multivariate Behavioral Research, 28,* 67-96.

Efron, B. (1982). *The Jackknife, the Bootstrap, and other Resampling Plans*. Philadelphia: SIAM.

Fleishman, A. I. (1978). A method for simulating nonnormal distributions. *Psychometrika, 43,* 521-532.

Fornell, C., & Rust, R. T. (1989). Incorporating prior theory in covariance structure analysis: A Bayesian approach. *Psychometrika, 54,*249-259.

Golden, R. M. (2000). Statistical tests for comparing possibly specified and non-nested models. *Journal of Mathematical Psychology, 44,* 153-170.

Kumar, A., & Sharma, S. (1999). A metric measure for direct comparison of competing models in covariance structure analysis. *Structural Equation Modeling, 6,* 169-197.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science, India, 12*, 49-55.

Mulaik, S. A. (2001). The curve-fitting problem: An objectivist view. *Philosophy of Science, 68*, 218-241.

Nevitt, J., & Hancock, G. R. (1999). NNORMULT and PWRCOEFF: A set of GAUSS programs for simulating multivariate nonnormal data. *Applied Psychological Measurement, 23*(1), 54.

Oczkowski, E., & Farrell, M. A. (1998). Discriminating between measurement scales using non-nested tests and two-stage least squares estimators: The case of market orientation. *International Journal of Research in Marketing, 15*, 349-366.

Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society, 37*, 81-91.

Rao, C. R. (1949). On the distance between two populations. *Sankhya, 9*, 246-248. *Journal of Mathematical Psychology, 44*, 153-170.

Rust, R. T., Lee, C., & Valente, Jr., E. (1995). Comparing covariance structure models: A general methodology. *International Journal of Research in Marketing, 12*, 279-291.

Rust, R. T., & Schmittlein, D. C. (1985). A Bayesian cross-validated likelihood method for comparing alternative specifications of quantitative models. *Marketing Science, 4*, 20-40.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

Smith, R. J. (1992). Non-nested tests for competing models estimated by generalized method of moments. *Econometrica, 60*, 973-980.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229-239.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika, 48*, 465-471.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica, 57*, 307-333.

Widaman, K. F. (1985). Hierarchically tested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1-26.

Wishart, J. (1928). Sampling errors in the theory of two factors. *British Journal of Psychology, 19*, 180-187.

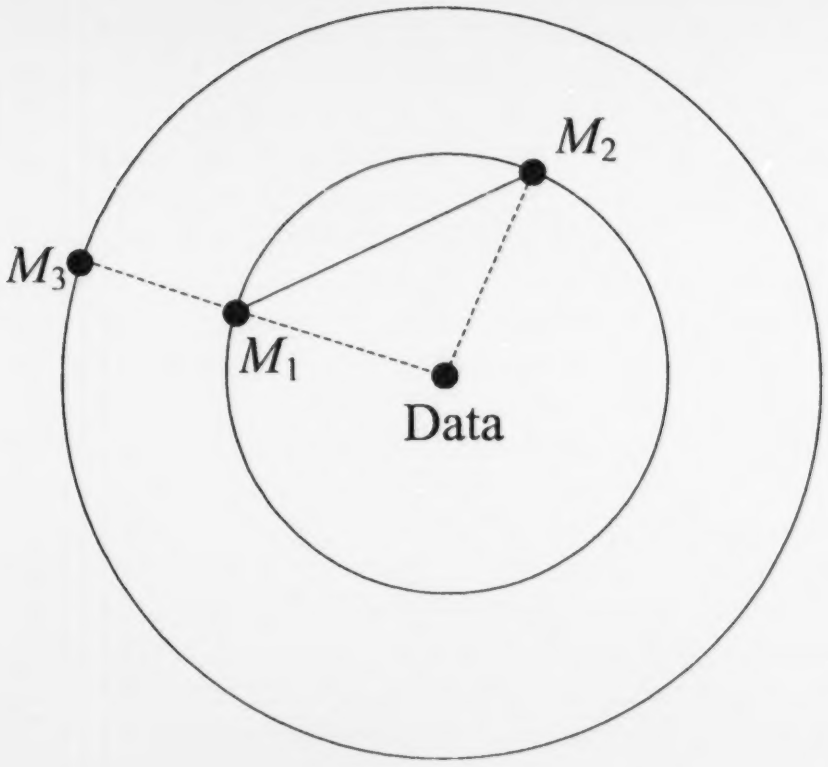

Figure 1: Dashed lines from models M_1, M_2, and M_3 to the data indicate data-model (mis)fit. The solid line indicates the discrepancy between M_1 to M_3, which cannot be assessed by indirect comparisons of the models' relations to the data.

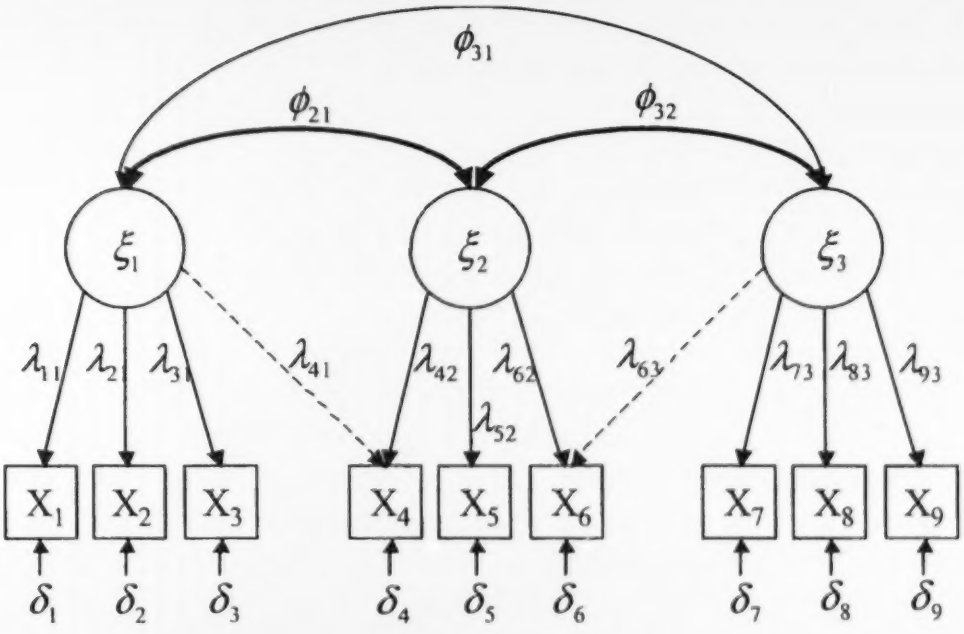

Figure 2: Model G is the model with all parameters present; Model 1 constrains λ_{41} and λ_{63} (the dashed paths) to be 0; Model 2 constrains ϕ_{21} and ϕ_{32} (the bold bi-directional factor covariances) to be 0; Model C imposes the constraints of both Model 1 and Model 2.

Measure	Mean	Standard Deviation	Minimum	Maximum
$r_{S\,1,C}$	0.998323	0.000959	0.993817	0.999433
$r_{S\,2,C}$	0.999477	0.000176	0.998764	0.999719
95^{th} %ile $RD_{1,2}$	0.203446	0.012445	0.175220	0.241889

Table 1: Model Comparison of nested and nonnested models

Appendix A

Instantiation of Model G for data generation, defined in terms of the matrices of loadings (Λ), factor variances and covariances (Φ), and error variances and covariances (Θ).

$$\Lambda = \begin{bmatrix} 1.2 & 0 & 0 \\ 1.1 & 0 & 0 \\ 0.7 & 0 & 0 \\ 0.8 & 0.6 & 0 \\ 0 & 1.1 & 0 \\ 0 & 0.6 & 1.1 \\ 0 & 0 & 0.9 \\ 0 & 0 & 0.8 \\ 0 & 0 & 0.9 \end{bmatrix} \qquad \Phi = \begin{bmatrix} 1.0 & 0.4 & 0.7 \\ 0.4 & 1.0 & 0.7 \\ 0.7 & 0.4 & 1.0 \end{bmatrix} \qquad \Theta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Appendix B

GAUSS program to compute Rao's Distance between covariance matrices 'a' and 'b'.

```
new;
a={1.0 0.1,
   0.1 1.0};
b={1.0 0.2,
   0.2 1.0};
c=a*inv(b);
evunsort=eig(c);
ev=rev(sortc(evunsort, 1));
lo=ln(ev);
mult=lo'*lo;
f=.5*mult;
rd=SQRT(f);
PRINT "Distance =" rd;
end;
```

It should be noted that there are other ways to compute *RD*. For example, Kumar and Sharma (1999) include a program based on computing eigenvectors. The program detailed here is used because it (1) is more in line with eqs. (4) and (5), and (2) is not based on eigenvectors, which are not unique, the use of which may result in different software packages producing different values for *RD*.

TM034958

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A Statistical Test For Nonnested Covariance Structure Models	
Author(s): Roy Levy and Gregory R. Hancock	
Corporate Source: University of Maryland	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature:	Printed Name/Position/Title: Roy Levy	
Organization/Address: 1230 Benjamin Building U of MD College Park MD 20742	Telephone: (301) 405-3623	FAX:
	E-Mail Address: levyr@wam.umd.edu	Date: 5/16/03

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Assessment and Evaluation
University of Maryland, College Park
1129 Shriver Lab
College Park, MD 20742